Analysis Data Groups and Statistical Analysis Tool

(6/29/2020)

IMG has developed a new tool to analyze isolate genomes, scaffold sets and metagenome bins or mixed groups of these entities within the IMG/MER system. In order to perform such analyses, a user will need to first create one or more Analysis Data Groups (ADGs).

(Dev site: https://img-dev.jgi.doe.gov/cgi-bin/ken/main.cgi?section=WorkspaceAdg)

From the **Workspace** menu item, there is a new submenu **Analysis Data Group**.

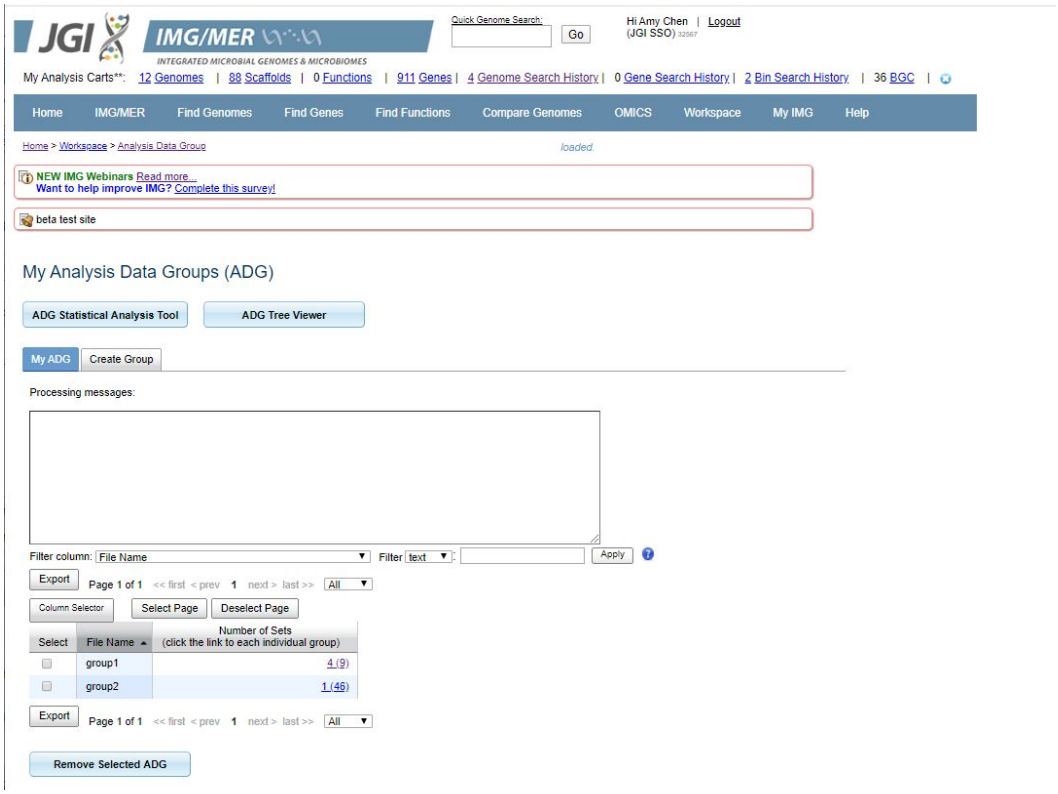


Figure 1: Analysis Data Groups (ADGs) main page

Use Case Scenario for statistical analysis using ADGs.

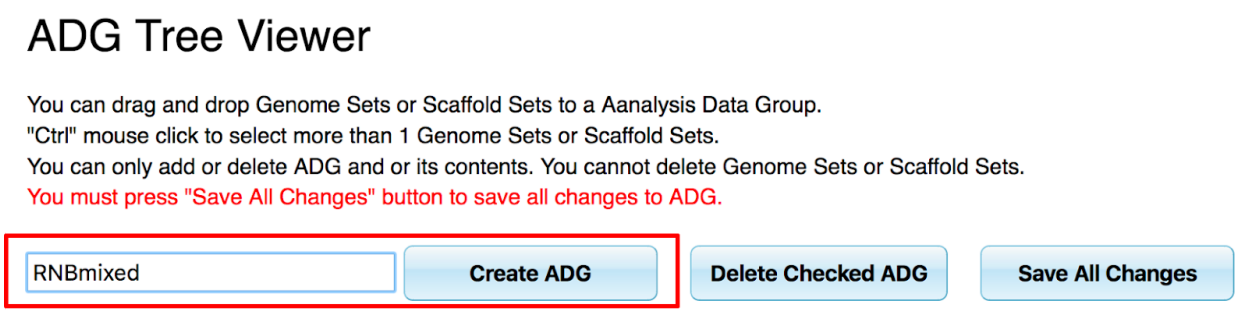
Example: Root Nodulating Bacteria (RNB) versus control group

Click ADG Tree Viewer. It is possible to create **mixed groups** containing isolate genomes AND metagenome-derived scaffold bins and treat these latter as "genome equivalents". In the example below - we compare TWO ADG sets ("RNBmixed" - composed of isolate genomes and scaffold bins of potential root nodulating bacteria VERSUS "RNBcontrol" - a control group containing all isolate genomes from non-plant-associated environments).

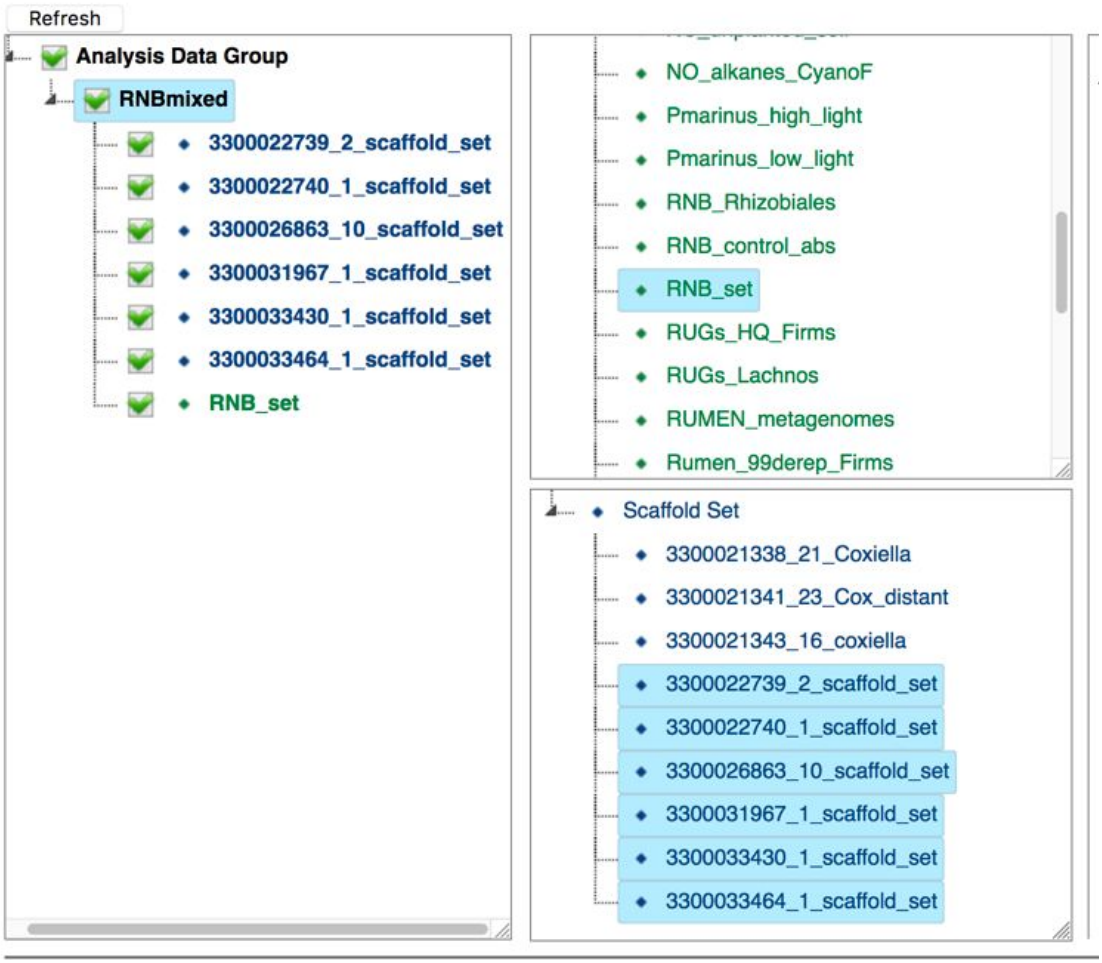
NOTE: ALL ADGs are created based on existing previously curated Genome Sets or Scaffold Sets

In this example, we create a mixed ADG (titled "RNBmixed") based on previously-curated sets of isolate genomes ("RNB Set" in Genome Set with 138 genomes) and SIX metagenome scaffold bins previously selected as potential root nodulating strains (3300022739_2_scaffold_set, 3300022740_1_scaffold_set, 3300026863_10_scaffold_set, 3300031967_1_scaffold_set, 3300033430_1_scaffold_set, 3300033464_1_scaffold_set).

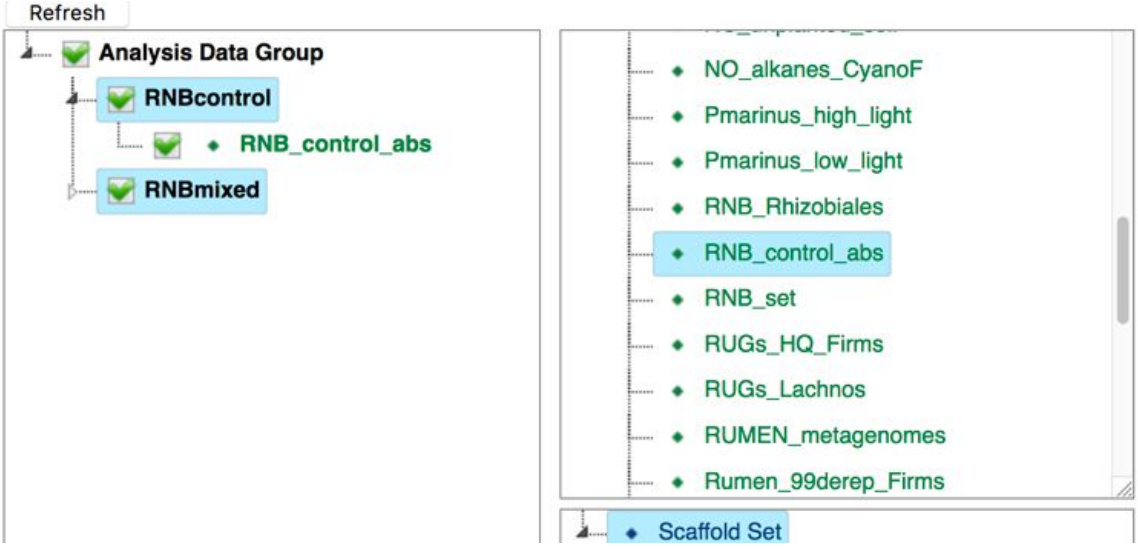
First, type in "RNBmixed" and click **Create ADG** button:



Newly created ADG appears in the window below. Simply click and drag the sets you wish to compile into this newly created ADG. In this example I choose "RNB_Set" (with 160 isolate genomes) from my pre-existing Genome Sets and 6 scaffold bin sets of putative RNB (previously identified based on the presence of known biomarkers of nodulation). **IMPORTANT NOTE:** click **Save All Changes** button. Expand "RNBmixed" to see added sets:



You will have to repeat the same steps as above to create the second control ADG set containing all isolates - create “RNBcontrol” ADG and click and drag “RNB_control” set from Genome Sets:



Select the two newly created ADG - and you are ready to proceed with the statistical comparisons. Proceed as directed in the stats tool guide.
138 isolate genomes + 6 scaffold bins (treated as genome “equivalents”) = 144 genomes in the “RNBmixed” ADG will be compared against the 82 isolate genomes in the “RNBcontrol” ADG.

Please watch our IMG webinar on using the regular Statistical Analysis tool for details.

Results table: Means of gene counts in RNBcontrol (n=82) and RNBmixed (n=144) are highlighted:

Download Full Results

Filter column: MWTest adjPval Filter text : Apply

Export Page 1 of 11 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Feature	Description	Mean RNBcontrol(n=82)	Mean RNBmixed(n=144)	MWTest adjPval
	KO:K14666	N-acetylglucosaminyltransferase [EC:2.4.1.-]	0	1.02778	9.691e-40
	KO:K14659	chitooligosaccharide deacetylase [EC:3.5.1.-]	0	0.965278	5.829e-37
	KO:K14658	nodulation protein A [EC:2.3.1.-]	0	1.15278	1.288e-36
	KO:K09695	lipooligosaccharide transport system ATP-binding protein	0.0609756	1.08333	1.528e-34
	KO:K09694	lipooligosaccharide transport system permease protein	0.0609756	1.09028	7.675e-34
	KO:K14657	LysR family transcriptional regulator, nod-box dependent transcriptional activator	0	2.13194	2.269e-32
	KO:K02592	nitrogenase molybdenum-iron protein NifN	0.231707	0.986111	4.637e-28
	KO:K13653	AraC family transcriptional regulator	0.0731707	0.875	1.135e-27
	KO:K02587	nitrogenase molybdenum-cofactor synthesis protein NifE	0.231707	1.01389	1.480e-27
	KO:K03855	ferredoxin like protein	0.219512	0.986111	1.982e-27
	KO:K00313	electron transfer flavoprotein-quinone oxidoreductase [EC:1.5.5.-]	0.219512	1.05556	2.014e-27
	KO:K02585	nitrogen fixation protein NifB	0.292683	1	8.841e-26